CSCI3160: Special Exercise Set 5

Prepared by Yufei Tao

Problem 1. Consider the alphabet Σ with letters a, b, c, d, e, f, g, h whose frequencies are 3%, 5%, 10%, 12%, 14% 16%, 18%, and 22%, respectively. Use Huffman's algorithm to find a prefix code on Σ that has the smallest average length.

Problem 2. Consider an alphabet Σ that contains n letters with their frequencies given, where n is a power of 2. Prove: the prefix code constructed using Huffman's algorithm has an average length of *at most* $\lceil \log_2 n \rceil$.

Problem 3. Describe how to implement Huffman's algorithm to ensure a worst-case time complexity of $O(n \log n)$, where n is the size of the alphabet Σ.

Problem 4. Consider the alphabet $\Sigma = \{1, 2, ..., n\}$ for some integer $n \geq 1$. Suppose that the frequency of i is *strictly higher than* the frequency of $i + 1$, for any $i \in [1, n - 1]$. Prove: in an optimal prefix code, for any $i \in [1, n - 1]$, the codeword of i cannot be longer than that of $i + 1$.

Problem 5. Consider an alphabet Σ with n letters, all of which have exactly the same frequency. The value of n is a power of 2. If we use Huffman's algorithm to generate the codewords for all the letters in Σ, how many bits are there in the shortest codeword?